



**HAL**  
open science

# The European intraday electricity market : a modeling based on the Hawkes process

Benjamin Favetto

► **To cite this version:**

Benjamin Favetto. The European intraday electricity market : a modeling based on the Hawkes process. 2019. hal-02089289

**HAL Id: hal-02089289**

**<https://hal.science/hal-02089289>**

Preprint submitted on 16 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The European intraday electricity market : a modeling based on the Hawkes process

Benjamin Favetto<sup>1</sup>

---

## Abstract

This article deals with the modeling of the trading activity on the European electricity intraday market by a self-exciting point process (also known as Hawkes process). It gives some empirical evidence of self-excitement, and discuss the time-homogeneity of the baseline of the process. The question of the functional shape of the intensity kernel is also adressed. Finally, a parameter estimation procedure is derived for the model with a non-constant baseline.

*Keywords:* European electricity intraday market, Self-exciting point process, change-point detection, parameter estimation

*2010 MSC:* 00-01, 99-00

---

## Introduction

Although electricity markets are quite different from other financial markets (due to strong technical constraints, and a few number of companies acting on them), observations of the intraday electricity market suggest that there exists  
5 clusters of operations during the trading period. Our aim is to give empirical evidence of this phenomenon, and study the statistical properties of a model based on a self-exciting process process for the dates of realized trades.

Indeed, the trading activity on intraday electricity market has significantly increased over the last years (Kiesel & Paraschiv (2017)). This may be driven  
10 by in increasing connectivity of networks, a diversification of sources (e.g. re-

---

\*INSEE, Département des Études Économiques  
*Email address:* `benjamin.favetto@insee.fr` (Benjamin Favetto)

newable ones), and an evolution of the demand. Therefore, it shed light on the necessity of improving the knowledge of the intraday electricity market, in order to give more interest on endogeneous mechanisms which can also drive the market activity.

15 From this point of view, it is important to be able to compare the market characteristics for different hours of the day, and eventually to detect changes in these characteristics on a long timescale (e.g. one year).

The article of Becker et al. (2013) models the price events on the Australian market, and focuses on the spot price spikes. The model is based on a self-  
20 exciting process, also known as Hawkes process. This class of point processes allows the intensity of jumps to depend on previous ones, as an extension of the Poisson process.

Although it is based on a self-exciting process, the framework of this article is quite different from what we want to study : first of all, it deals with the  
25 price process and looks for exogenous variables which could explain variations in the intensity of price spikes, then one of its aims is to identify seasonality and patterns at a weekly and daily level. Our study focus on the dates of trades, to give an evidence of the influence of past trades on the current activity, and some statistical properties of the *memory* of such a process.

30 Theoretical properties of the Hawkes process have been widely investigated, and it is of major interest for financial modeling(Bacry et al. (2013a)), criminology (Lewis et al. (2012)) and earthquake analysis (Ogata (1988)). Moreover, several asymptotic results lead to derive estimating procedures for high frequency financial data (Da Fonseca & Zaatour (2014), Da Fonseca & Zaatour  
35 (2015), Bacry & Muzy (2014)) and genome analysis (Reynaud-Bouret et al. (2010)). However, and even if the Hawkes process appears to be an efficient natural extension of the Poisson process, few models based on its properties can be found in the field of energy econometrics (Becker et al. (2013) is the main one).

40 Then, our goal is to justify the interest of taking into account trade clustering in the intraday market, with a modeling based on a point process for the dates

of trades. We mainly focus on the dates of the realized trades in a given period of the day, in order to try to enlighten the specific features of this market.

This article presents the goals of the study, the theoretical issues, and the  
45 numerical results on simulated and real datasets.

## 1. The Hawkes process : a tractable self-exciting point process

Some properties of the univariate Hawkes process are briefly recalled here.

**Definition 1.** *Let  $(\Omega, \mathcal{F}, \mathbf{P})$  a probability space. A Hawkes process  $X_t = (\lambda(t), N(t))$  is a stochastic process on the state space  $\mathcal{D} = \mathbb{R}_+ \times \mathbb{N}$ , where  $N(\cdot)$  is a simple point process adapted to  $\mathcal{F}$  determined by its intensity process  $\lambda(\cdot)$ :*

$$\begin{aligned} \mathbf{P}(N(t+h) - N(t) = 1 | \mathcal{F}_t) &= \lambda(t)h + o(h) \\ \mathbf{P}(N(t+h) - N(t) > 1 | \mathcal{F}_t) &= o(h) \\ \mathbf{P}(N(t+h) - N(t) = 0 | \mathcal{F}_t) &= 1 - \lambda(t)h + o(h) \end{aligned}$$

and  $\lambda(t)$  (given  $\mathcal{F}_t$ ) is defined by:

$$\lambda(t) = \mu(t) + \int_0^t \varphi(t-s) dN(s) \tag{1}$$

where  $\mu$  is a positive function called baseline and  $\varphi$  is a non-increasing and non-negative function called kernel.

50 The knowledge of  $\mu$  and  $\varphi$  determines  $\lambda(\cdot)$  and then the distribution of  $N(\cdot)$ . To give some details about the notations of Equation 1,  $\int_0^t \varphi(t-s) dN(s) = \sum_{t_i < t} \varphi(t-t_i)$  where  $\{t_i\}$  is the collection of ordered times of a unitary jump of  $N$ . Hence a jump of  $N(\cdot)$  at time  $t_i$  will increase the intensity for the dates  $t > t_i$  : this explains the use of the term " self-exciting process". Furthermore,  
55 the memory of former jumps depends on the tail of the kernel function.

. In the literature of the applications of self-exciting processes, two types of kernels are mostly considered :

- the exponential kernel (*i.e.*  $\varphi(t) = \alpha \exp(-\beta t) \mathbf{1}_{(0,+\infty)}(t)$  )
- the power-law kernel (*i.e.*  $\varphi(t) = \frac{\alpha}{(c+t)^\beta} \mathbf{1}_{(0,+\infty)}(t)$  )

60 For these kernels, the parameter  $\beta$  is the decay parameter : it determines the tail of the kernel function. Moreover, if  $\|\varphi\|_1 < 1$ , the Hawkes process admits a version with stationary increments.

## 2. The European intraday electricity market data

### 2.1. Description of the data

65 A database of intraday prices is available, based on the EPEX SPOT Market. EPEX SPOT market is an European exchange place for power spot trading in Germany, France, the United Kingdom, the Netherlands, Belgium, Austria, Switzerland and Luxembourg. It currently connects markets representing 85% of the European power consumption.

70 Intraday markets, which are organized by continuous trading-orders of the members, are entered perpetually into the order book. As soon as two orders are compatible, the trade is executed. Contracts for hourly quantities of power can be traded up to 30 minutes before physical fulfillment.

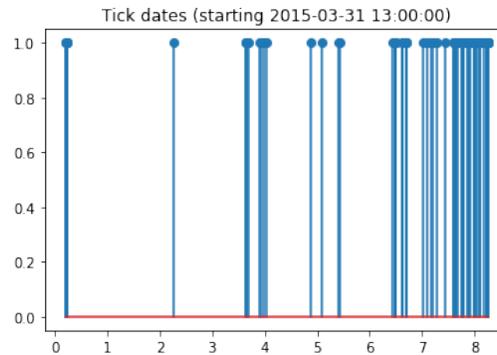


Figure 1: Timestamps from the database (executed trades)

Figure 1 represents the tick dates on the electricity spot market, for a given

75 delivery date (2015-03-31 22:00:00 UTC<sup>1</sup>), with trading start at 2015-03-31  
 13:00:00 UTC<sup>2</sup> and trading end at 2015-03-31 21:15:00 UTC. (At this date,  
 contracts for hourly quantities could be traded up to 45 minutes before fulfill-  
 ment. Now, they can be traded up to 30 minutes before fulfillment.)

2.2. Descriptive statistics: empirical evidence of trade clustering

80 Some justifications of the use of Hawkes process as tool for modeling the  
 intraday electricity market trading activity are presented in this section. First,  
 inter-arrival times of trades are studied to shed light on clustering phenomenon,  
 as Da Fonseca & Zaatour (2014) did for the stock market.

Figure 2 shows that the QQ-plot of inter-arrival times of trades against an  
 85 exponential distribution and rejects clearly the homogeneous Poisson-process as  
 a data-generating process for the order flow.

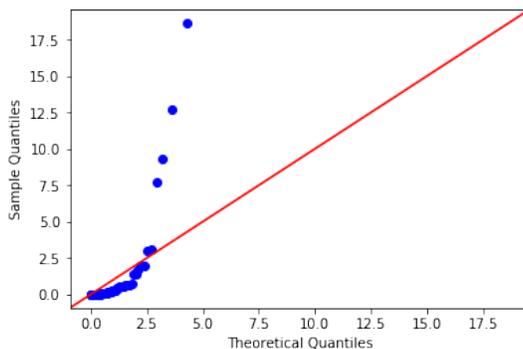


Figure 2: QQ-plot of inter-trade durations against exponential distribution

In addition, a self-exciting point process shows positive covariance between  
 collections of events in time (Lewis et al. (2012)) : for  $t_1 < t_2 < t_3$ ,

$$\text{Cov}(N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}) > 0$$

<sup>1</sup>This correspond to the hour 00:00:00 of 2015-04-01 in France and Germany.

<sup>2</sup>Add two hours for the time in the countries which act on the intraday market.

8.55e-01	-3.45e-02	1.24e-01	-6.90e-03	1.20e+00	-1.31e-01	2.48e-01	2.55e-01	1.72e-01	3.06e+00
-3.45e-02	1.44e-01	-5.75e-03	5.75e-03	-8.05e-02	5.75e-02	-2.13e-01	-2.41e-01	-4.02e-02	8.39e-01
1.24e-01	-5.75e-03	2.54e-01	8.39e-02	4.46e-01	6.46e-01	2.26e-01	-1.79e-01	2.82e-01	-1.77e-01
-6.90e-03	5.75e-03	8.39e-02	7.23e-01	6.37e-01	6.44e-02	-6.78e-02	3.93e-01	1.24e+00	-6.92e-01
1.20e+00	-8.05e-02	4.46e-01	6.37e-01	3.15e+00	9.89e-02	7.77e-01	1.19e+00	2.05e+00	5.89e+00
-1.31e-01	5.75e-02	6.46e-01	6.44e-02	9.89e-02	6.40e+00	2.49e+00	1.06e+00	2.32e+00	-2.84e+00
2.48e-01	-2.13e-01	2.26e-01	-6.78e-02	7.77e-01	2.49e+00	5.55e+00	4.23e+00	4.45e+00	-6.87e-01
2.55e-01	-2.41e-01	-1.79e-01	3.93e-01	1.19e+00	1.06e+00	4.23e+00	1.17e+01	8.48e+00	3.37e+00
1.72e-01	-4.02e-02	2.82e-01	1.24e+00	2.05e+00	2.32e+00	4.45e+00	8.48e+00	2.98e+01	7.30e+00
3.06e+00	8.39e-01	-1.77e-01	-6.92e-01	5.89e+00	-2.84e+00	-6.87e-01	3.37e+00	7.30e+00	1.72e+02

Table 1: Empirical covariance of the number of transactions  $N((k+1)\Delta) - N(k\Delta)$

(A close formula for this covariance can be found in Bacry et al. (2012) for the stationary case, and in Proposition 3 for the general case). Conversely, for a Poisson process, independent increments entail that this covariance vanishes.

90 Table 1 gives the empirical covariance of  $N((k+1)\Delta) - N(k\Delta)$ ,  $0 \leq k < 10$ , for  $\Delta = 0.1$ , computed for the first hour of the day with 30 observations in April 2015. Most of the values are significantly positive.

Furthermore, the histogram of the dates (Figure 3) underpins time-inhomogeneity, as the number of ticks per bin of the histogram strongly varies over the time.

95 This variation can be explained by :

- Time-inhomogeneity of the baseline (*i.e.*  $\mu$  depends on  $t$  and the underlying model is an inhomogeneous Poisson-process),
- Self-excitement phenomenon.

100 As an univariate Hawkes-process based model can recover both sources of variation, a model based on such a process is chosen for the dates of executed trades on the intraday electricity market for a given hour. In addition, a model based on a Hawkes process can handle data generated by a Poisson process, considering  $\varphi$  as the null function. Then, the models are embedded.

### 3. Theoretical properties of self-exciting point processes

105 We deal with  $n$  independent and identically distributed sets of observations  $N_1(\cdot), \dots, N_n(\cdot)$  of a data-generating process  $N$  which is assumed to be a Hawkes

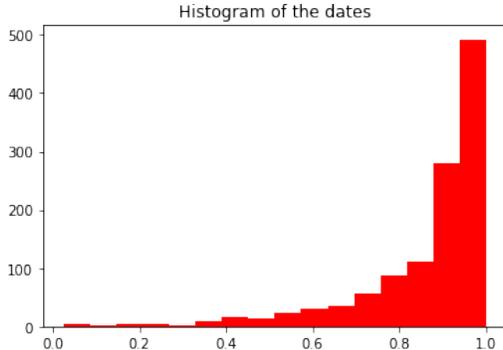


Figure 3: Distribution of the dates of executed trades (April 2015)

process observed over a finite time interval ( $[0, 1]$  for sake of simplicity). This corresponds to  $n$  observations of the trades concerning a trading period related to a fixed hour of the day, during consecutive trading days. We assume that these observations are i.i.d., even if seasonality, weather or position of the day in the week could be taken into account in a more refined model. This choice is not the most common one : several seminal papers deal with a unique stationary process observed on a long time range, but we can not assume that for one observed period, because of the low number of trades during the trading period<sup>3</sup> on the intraday electricity market.

As the length of the time interval is fixed<sup>4</sup>, for the theoretical properties of the estimation we consider observations on  $[0, 1]$  (which has no impact, up to a rescaling). Moreover, we can observe a sample of  $n$  days, assumed independent and identically distributed (typically  $n \sim 30$  to deal with one month observations). This determines the choice of our asymptotic framework.

Hence, we have :

- $n$  sets of ordered dates  $\{t_j^i\}$ , observations of independent point process  $N_i$

<sup>3</sup>  $\sim 8$  hours for the first hour of the day, and up to more than 20 hours

<sup>4</sup>The trading period for hourly quantities had a duration of 8 hours 15 minutes in 2015, and 8 hours 30 minutes now, for the first hour of the day. As all trading periods start at 15pm, the results can be impacted. Then we mainly focus on the first hour of the day in numerical applications.

( $1 \leq i \leq n$ ) over  $[0, 1]$ ,

- the common baseline intensity  $\mu(t)$ , which may vary during the trading period,
- the common kernel is  $\varphi$  (non-negative and non-increasing function)

The intensity of  $N_i(t)$  (given the information  $\mathcal{F}_t$  at time  $t$ ) is :

$$\lambda_i(t) = \mu(t) + \int_0^t \varphi(t-s) dN_i(s) \quad \text{for } t \in [0, 1]$$

As an increasing number of ticks can be observed as  $t \rightarrow 1$ , we may test if  $\mu$  could be considered as a constant or an increasing function. We insist on the fact that the "memory size" (determined by the decay parameter) is of major interest. As a property of the intraday market, it gives a better understanding of the self-exciting property, in comparison with a Poisson process.

**Remark 1.** *Even if the exponential and the power law kernel have an unbounded support, their mass is concentrated on a small interval when the decay parameter is large.*

### 3.1. Covariance of the Hawkes process

We recall here some properties of the Hawkes process, in the general context of a time-varying baseline function. Some of the proofs are gathered at the end of the report in the Appendix.

**Proposition 1.** *Let  $M(t) = N(t) - \int_0^t \lambda(s) ds$ . Then*

- $\{M(t), 0 \leq t \leq 1\}$  is a zero-mean square integrable martingale with respect to the filtration  $(\mathcal{F}_t)$
- $d[M]_t = dN_t$ , where  $[.]$  is the quadratic variation of the martingale (i.e.  $[M]_t = \sum_{s \leq t} (M(s) - M(s-))^2$ )
- $N(t) = M(t) + \int_0^t \mu(s) ds + \int_0^t \int_0^s \varphi(s-u) dN(u) ds$  for  $t \in [0, 1]$

(See, for instance, Chen & Hall (2013).)

. Here is recalled a simple but important result for the further asymptotics. The mean of the process can be expressed as the solution of an integral equation.

**Lemma 1.** (*Integration by parts and convolution formula*)

Let  $m(t) = \int_0^t \mu(s)ds$ . For  $t \in [0, 1]$  :

•

$$\int_0^t \int_0^s \varphi(s-u) dN(u) ds = \int_0^t \varphi(t-s) N(s) ds$$

•

$$H(t) = \mathbf{E}(N(t)) = m(t) + \int_0^t \varphi(t-s) \mathbf{E}(N(s)) ds$$

150 (See Lemma 2 in Bacry et al. (2013b).)

. Let define  $\varphi_n = \varphi \star \dots \star \varphi$  ( $n$  times -convolution product) and  $\psi = \sum_{n \geq 1} \varphi_n$ . The idea of using integral equations and convolution product leads to useful and tractable computations, as in Bacry et al. (2012) or Bacry et al. (2013b). Notice that the key quantities of the distribution of  $N(t)$  are expressed with  $\mu$ ,  $\psi$  and  
155  $M$ .

**Proposition 2.** (*Formulas based on the convolution product*)

For  $t \in [0, 1]$  :

•

$$H(t) = m(t) + \int_0^t \psi(t-s) m(s) ds = m(t) + (\psi \star m)(t)$$

•

$$N(t) = m(t) + (\psi \star m)(t) + M(t) + (\psi \star M)(t)$$

•

$$\lambda(t) = \mu(t) + (\psi \star \mu)(t) + (\psi \star dM)(t)$$

$$\text{with } (\psi \star dM)(t) = \int_0^t \psi(t-s) dM(s).$$

(See Lemma 4 in Bacry et al. (2013b).)

160 The covariance of the process (on which some estimators are build, see Bacry  
 et al. (2012) and Bacry & Muzy (2016)) can be computed in the non-stationary  
 case.

**Proposition 3.** For  $t_1 < t_2$  and  $t_3 < t_4$ ,

$$\begin{aligned} \text{Cov}(N(t_2) - N(t_1), N(t_4) - N(t_3)) &= \mathbf{E} \left( \int_{t_1}^{t_2} \int_{t_3}^{t_4} dN(s)dN(u) \right) - \int_{t_1}^{t_2} \int_{t_3}^{t_4} h(s)h(u)dsdu \\ &= \int_{t_1 \vee t_3}^{t_2 \wedge t_4} h(s)ds + \int_{t_1}^{t_2} \int_{t_3}^{t_4} \gamma(s, u)dsdu \end{aligned}$$

with

$$\gamma(s, u) = \psi(s - u)h(u) + \psi(u - s)h(s) + \int_0^{s \wedge u} \psi(s - v)\psi(u - v)h(v)dv$$

The last formula ensures that, in the case of a self-exciting process, the co-  
 165 variance of the number of jumps during two non-overlapping intervals is positive.  
 However, building an estimator based on deconvolution with this quantity seems  
 to be untractable. The interest is theoretical yet : it validates our numerical  
 justification of self-excitation (Table 1).

### 3.2. Parameter estimation : maximum likelihood estimators

170 This subsection presents the asymptotic results for the maximum likelihood  
 estimator in both asymptotic frameworks, under classical regularity assump-  
 tions. We will see in Section 6 that, for our data, supplementary material is  
 needed, because of the occurrence of a change-point in the baseline, but the ba-  
 sis results are essential to provide convergence rate and asymptotic confidence  
 175 intervals.

As the log-likelihood of a Hawkes process has a closed form, up to the spec-  
 ification of a functional form for  $\mu(\cdot; \theta)$  and  $\varphi(\cdot; \theta)$ , it is possible to perform  
 maximum likelihood estimation.

Let  $(X_t)$  be an univariate simple point process on  $[0, 1]$  and  $0 \leq t_1 < \dots <$   
 $t_N \leq 1$  denote an observation set of ordered dates of jumps of  $(N_t)$ . Then

the log-likelihood of  $(X_t)$  (written, up to a constant, with respect to a Poisson process measure of intensity 1, Ogata (1988)) is:

$$\ell = \int_0^1 \ln(\lambda(s)) dN_s - \int_0^1 \lambda(s) ds \quad (2)$$

Then

$$\ell = \sum_{t_i} \ln \left( \mu(t_i) + \sum_{t_j < t_i} \varphi(t_i - t_j) \right) - \int_0^1 \mu(s) ds - \sum_{t_i} \int_{t_i}^1 \varphi(t - t_i) dt$$

From this expression, a maximum-likelihood estimator (MLE) can be derived (Ogata (1988)).

Assuming that  $n$  i.i.d Hawkes processes  $N_1, \dots, N_n$  are observed (with respective intensities  $\lambda_i(t)$ ) over  $[0, 1]$ , the log-likelihood is:

$$\ell_n(\theta) = \sum_{i=1}^n \ell^{(i)}(\theta)$$

where  $\ell^{(i)}(\theta)$  is the log-likelihood associated with the set of observations  $\{N_i(t), t \in [0, 1]\}$ . Then

$$\ell_n(\theta) = \sum_{i=1}^n \left( \int_0^1 \ln(\lambda_i(t; \theta)) dN_i(t) - \int_0^1 \lambda_i(t; \theta) dt \right) \quad (3)$$

**Definition 2.** *The score function is:*

$$S_n(\theta) = \partial \ell_n(\theta) = \sum_{i=1}^n \left( \int_0^1 \frac{\partial_\theta \lambda_i(t; \theta)}{\lambda_i(t; \theta)} dM_i(t; \theta) \right) \quad (4)$$

where  $dM_i(t; \theta) = dN_i(t) - \lambda_i(t; \theta)$ .

First, denoting  $H(t; \theta_0) = \mathbf{E}_{\theta_0}(N(t))$ , the uniform law of large numbers holds (Proposition 7) :

$$\sup_{t \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n N_i(t) - H(t; \theta_0) \right| \xrightarrow{\mathbf{P}} 0$$

Then, the score function converges at rate  $n$ :

$$\frac{1}{n}S_n(\theta) \xrightarrow{\mathbf{P}} s(\theta, \theta_0) = \mathbf{E}_{\theta_0} \left( \int_0^1 \frac{\partial_\theta \lambda(t; \theta)}{\lambda(t; \theta)} \{dN(t) - \lambda(t, \theta)dt\} \right)$$

and the limit function only depends on the intensity of the process:

$$s(\theta, \theta_0) = \mathbf{E}_{\theta_0} \left( \int_0^1 \frac{\partial_\theta \lambda(t; \theta)}{\lambda(t; \theta)} (\lambda(t; \theta_0) - \lambda(t; \theta))dt \right)$$

This limit is slightly different from those of Chen & Hall (2013). Indeed, in their framework,  $\lambda(t, \theta)$  is replaced by its expectation. Moreover, the MLE is asymptotically Gaussian, with a different asymptotic variance.

**Theorem 1.** *Under the following conditions :*

(C1)  $\mu(\cdot; \theta)$  and  $\varphi(\cdot; \theta)$  are positive and continuous on  $[0, 1]$  for all  $\theta \in \Theta$  ;

(C2) The parameter space  $\Theta$  is compact and its interior is connected and contains a  $d$ -dimensional non empty open ball which contains the true parameter ;

190 (C3) For each  $t \in [0, 1]$ , the functions  $\mu(t; \theta)$  and  $\varphi(t; \theta)$  are twice continuously differentiable in  $\theta$  and their partial derivatives up to order 2 with respect to  $\theta$  are uniformly equicontinuous when regarded as families of functions of  $\theta$  ;

(C4) For each  $\theta$ ,  $\partial_\theta \varphi(t, \theta)$  and  $\partial_{\theta\theta'}^2 \varphi(t; \theta)$  are continuously differentiable in  $t$ .

(C5) The matrix-valued function :

$$\gamma(\theta) = \int_0^1 \frac{\{\partial_\theta \mu(t; \theta) + \int_0^t \partial_\theta \varphi(t-s; \theta)h(u; \theta)du\}^{\otimes 2}}{\mu(t; \theta) + \int_0^t \varphi(t-u; \theta)h(u; \theta)du} dt$$

195 is non singular at  $\theta_0$ .

As  $n \rightarrow +\infty$ , with probability tending to 1, the maximum likelihood estimator  $\hat{\theta}$  exists as a solution to the score equation  $S(\theta) = 0$  and  $\hat{\theta}$  tends to  $\theta_0$  in

probability. In addition,

$$\sqrt{n} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I(\theta_0)^{-1})$$

where the information matrix is defined by:

$$I(\theta_0) = \mathbf{E}_{\theta_0} \left( \int_0^1 \frac{(\partial_\theta \lambda(t, \theta_0))^2}{\lambda(t, \theta_0)} dt \right)$$

Moreover, when  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{k=1}^n \left\{ \int_0^1 \frac{\partial_\theta \lambda_k(t; \hat{\theta})^{\otimes 2}}{\lambda_k(t; \hat{\theta})^2} dN_k(t) - \int_0^1 \frac{\partial_\theta^2 \lambda_k(t; \hat{\theta})}{\lambda_k(t; \hat{\theta})} dM_k(t; \hat{\theta}) \right\} \xrightarrow{\mathbf{P}} I(\theta_0)$$

These results still require regularity assumptions on the baseline and kernel functions with respect to the parameter, therefore they cannot be used if there is an abrupt change-point in the baseline. But they are relevant both in theory and in practice : they give the rate of convergence for the MLE, and they provide asymptotic confidence regions for the parameters.

#### 4. Empirical evidence of a non-constant baseline modeling

In this section we propose a test to give the evidence of using a non-constant baseline in the model : the intensity of the point process is increasing over the time not only by the self-excitement part but also due to a change in the baseline. In the literature, few papers investigate this case. For instance, Rambaldi et al. (2018) deals for instance with intensity bursts, defined as a short time period during which the number of counts is larger than the typical count rate. This case is different from a permanent change in the intensity but suggests to use a parametric approach to estimate the model and an information criterion to discriminate whether there is a change-point or not.

At this point, a change-point estimation is performed as a preliminary work to the estimation of a self-exciting process including a non-constant baseline, in

order to justify a more complex modeling (with respect to the usual assumptions in the literature).

215 First, an empirical evidence of a non-constant baseline is given from the data. Indeed, in this section, a procedure is proposed to detect a change-point in the baseline, and to estimate its date and its magnitude, without any assumption on the shape of the kernel function. This aims to justify to deal with non-stationarity for the data-generating process.

220 *4.1. Detection of a changepoint based on a kernel estimation*

The detection of a rough variation in the baseline function is based on a kernel estimation.

**Proposition 4.** *For  $t \in [0, 1]$ , let  $\mathcal{N}_n(t) = \sum_{i=1}^n N_i(t)$  be the sum-process, and  $\Lambda_n(t) = \sum_{i=1}^n \lambda_i(t)$ . The process  $(\mathcal{N}_n(t))$  is a Hawkes process on  $[0, 1]$ , with intensity  $(\Lambda_n(t))$ . Moreover,*

$$\Lambda_n(t) = n\mu(t) + \int_0^t \varphi(t-s)d\mathcal{N}_n(s)$$

**Remark 2.** *From a theoretical point of view, a single point process can be considered, with a baseline intensity depending on  $n$ . This is precisely a particular*  
 225 *case of the asymptotic assumption described in Chen & Hall (2013), Section 3, for the consistency and the asymptotic normality of the Maximum Likelihood Estimator (MLE) for a non-stationary Hawkes process. In this case, the intensity is increasing to infinity, which increases the number of points over  $[0, 1]$ .*

**Lemma 2.** *For  $t \in [0, 1]$ , let*

$$h(t) = \mu(t) + \int_0^t \psi(t-s)\mu(s)ds$$

*Then  $H'(t) = h(t)$ .*

**Proposition 5.** *For every continuous function  $f$  defined over  $[0, 1]$ , the following convergence in probability holds:*

$$\frac{1}{n} \int_0^1 f(t) d\mathcal{N}(t) \xrightarrow{\mathbf{P}} \int_0^1 f(t) h(t) dt$$

A kernel estimator of  $h$  can be derived from this proposition. Hence, let  $K$  a positive symmetric bounded kernel and  $b > 0$  a bandwidth and define for  $t \in [0, 1]$

$$\hat{h}(t) = \frac{1}{n} \int_0^1 \frac{1}{b} K\left(\frac{t-s}{b}\right) d\mathcal{N}(s) \quad (5)$$

230 The convergence of our estimator is granted by the following result of convergence for the mean-square error of a kernel-estimator.

**Proposition 6.** *(Mean Square Error)*

Assume that  $\int u^2 K(u) du < \infty$  and  $K$  has a compact support. If the number of observations  $n \rightarrow +\infty$ , and the bandwidth  $b = b_n \rightarrow 0$  with  $nb_n \rightarrow \infty$ , then there exists a positive constant  $C$  such that, for all  $t \in (0, 1)$ ,

$$\mathbf{E} \left( \left( \hat{h}(t) - h(t) \right)^2 \right) \leq C \left( \frac{1}{nb_n} + b_n^4 \right)$$

Hence, the behavior of  $h$  can be easily recovered by  $\hat{h}$ , except if a change-point occurs close to 0 or 1, due to side effects.

235 **Remark 3.** *Furthermore, an optimal  $b_n$  can be derived from the last inequality: taking  $b_n \propto n^{-1/5}$  ensures that the MSE is minimal (up to a constant). In this case, the MSE is of order  $n^{-4/5}$ .*

. On a simulation study, the asymptotic behaviors of  $n^{-1}\mathcal{N}(t)$  with different baselines (constant, piecewise affine and piecewise constant) are presented (Figure 4).  
240

For the sake of simplicity, and because we are looking for a decision whether  $\mu$  could be considered as constant (which is the case for stationary processes)

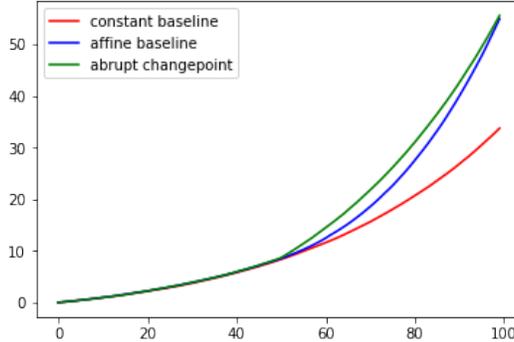


Figure 4: Asymptotic behaviour of  $n^{-1}\mathcal{N}(t)$  with different baselines

or not, we choose to consider only the piecewise constant case. Indeed, few  
 difference can be observed in  $h$  between a piecewise constant and a piecewise  
 245 affine baseline, depending on the values of the parameters (which can rely on  
 identification problems). Moreover, a piecewise constant baseline leads to easier  
 computations.

**Remark 4.** *In Chen & Hall (2016), a kernel estimator of  $h$  is introduced, while  
 the kernel  $\varphi$  is specified up to a parameter. However, our problem addresses the  
 250 question of the distinction of exponential and power-law kernel, that justifies to  
 look forward more investigations about the kernel estimation.*

We focus from now on the following test problem to detect if the baseline  
 might be considered as constant over the time or not :

$$\mathcal{H}_0 \quad \mu(t) = \mu_0 \text{ for } t \in [0, 1]$$

$$255 \mathcal{H}_1(\delta) \quad \mu(t) = \mu_0 + (\mu_1 - \mu_0)\mathbf{1}_{\{t \geq \delta\}} \text{ with } \mu_1 > \mu_0$$

Hence, rejecting  $\mathcal{H}_0$  for  $\mathcal{H}_1(\delta)$  means that there exists a date  $\delta \in (0, 1)$  such  
 that the baseline is piecewise constant on  $[0, \delta)$  and  $[\delta, 1]$  and increases with a  
 step of size  $\mu_1 - \mu_0$ . The following lemma gives details about the regularity of  
 $h$  when a change-point occurs.

**Lemma 3.** Under  $\mathcal{H}_1(\delta)$ , for  $t \in [0, 1]$

$$h(t) = \mu_0 \left( 1 + \int_0^t \psi(s) ds \right) + (\mu_1 - \mu_0) \left( 1 + \int_0^{t-\delta} \psi(s) ds \right) \mathbf{1}_{\{t \geq \delta\}}$$

260 We can perform a simulation study and underline that an abrupt change in  $\hat{h}$  relies on change point detection (there is a discontinuity point). Hence the values of  $\delta$  and  $\mu_1 - \mu_0$  can be estimated.

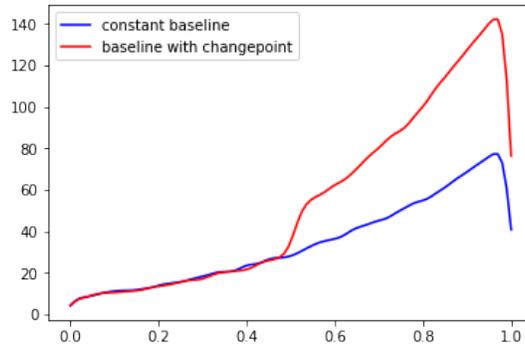


Figure 5: Changepoint detection with kernel estimation (simulation study)

This method is applied to the market data, first with a graphical approach. In Figure 6, a kernel estimation of  $h$  is performed with a Gaussian kernel (even if, in the theory, results are given for a kernel with compact support). Due to side effects, values close to 0 and 1 cannot be taken into account. 265

A clear change at 80% of the trading period is observed, of order of magnitude  $\sim 200$ . This fact agrees with the hypothesis that the intensity of the point process increases at the end of the trading period, and not only because of the self exciting effect. 270

. To the end of this section, we give a more precise statement for the test performed on  $h$ . Let  $(j\Delta, j = 0, \dots, N)$  be a regular grid on  $[0, 1]$ . Under assumption  $\mathcal{H}_0$ ,  $h$  is a Lipschitz function of Lipschitz constant  $L$  (due to the regularity of  $\psi$ , inherited from those of  $\varphi$ ), then the difference  $h((j+1)\Delta) - h(j\Delta)$

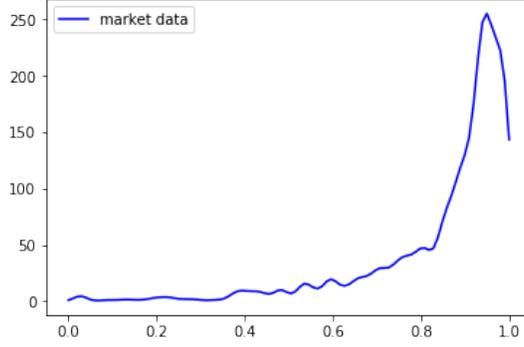


Figure 6: Change point detection based on a kernel estimation of  $h$  (market data, April 2015)

275 must be lower than  $L\Delta$ .

The following lemma states that, under  $\mathcal{H}_0$ , the observation of a big gap between two consecutive values of  $\hat{h}$  over the grid is an event of low probability.

**Lemma 4.** *Under  $\mathcal{H}_0$ , for  $\varepsilon > L\Delta$ , there exists a positive constant  $C$  (depending on the kernel  $K$ ) such that, for all  $0 \leq j \leq N - 1$ ,*

$$\mathbf{P}(|\hat{h}((j+1)\Delta) - \hat{h}(j\Delta)| \geq \varepsilon) \leq \frac{C}{n^{4/5}(\varepsilon - L\Delta)^2}$$

Up to the knowledge of  $C$  and  $L$ ,  $\varepsilon$  can be chosen such that  $\mathbf{P}(|\hat{h}((j+1)\Delta) - \hat{h}(j\Delta)| \geq \varepsilon) = \alpha$ , for a fixed  $\alpha \in (0, 1)$ . Then, if for a given  $j$ ,  $|\hat{h}((j+1)\Delta) - \hat{h}(j\Delta)| > \varepsilon$ , we can reject  $\mathcal{H}_0$  at level  $1 - \alpha$ . (Proof in the Appendix.)

**Example 1.** *Assume that  $\varphi(t) = \alpha e^{-\beta t}$  for  $t > 0$ . Then for  $k \geq 1$  and  $t > 0$  we have (by induction) :*

$$\varphi^{*k}(t) = \alpha^k \frac{t^{k-1}}{(k-1)!} e^{-\beta t}$$

and  $\psi(t) = \sum_{k \geq 1} \varphi^{*k}(t) = \alpha e^{-(\beta-\alpha)t}$ .

If we assume the condition  $\|\varphi\|_1 = \frac{\alpha}{\beta} < 1$ , then  $\beta - \alpha > 0$ . Moreover, with a constant baseline  $\mu_0$ ,  $h(t) = \mu_0(1 + \int_0^t \psi(s)ds)$  and  $L \leq \sup_{[0,1]} |h'(t)| = \mu_0\alpha$ .

#### 4.2. Estimation on different hours

285 The estimation procedures are performed on different hours of the day. The datasets come from April 2015, and consecutive days give, for each hour, 30 independent and identically distributed observations of a trajectory of the data-generating process. We may underline that the trading period varies for the different hours : it starts at 3pm the day before and ends 45 minutes before the  
290 hour of delivery.

Figure 7 presents the results of the non-parametric estimation of  $h$  for each hour : the estimator  $\hat{h}$  is strongly increasing in the last 20% of the period. We conclude that, for each hour, there may exist a change-point in the baseline at the end of the trading period.

### 295 5. Non-parametric estimation of the kernel based on splines

One major issue about the electricity intraday market is to make inference about the shape and the decay of the kernels, to have a preliminary result. Hence, this motivates a non-parametric approach, and estimate  $\varphi$  as a function. Once the shape of the kernel is guessed, a parameter estimation would give  
300 more accurate results (and we expect to derive the asymptotic distribution of the estimator under mild assumptions).

Other applications of the Hawkes process have discussed these types of kernel. Bacry et al. (2012) find a slowly decaying kernel shape (heavy-tailed power-law kernel) in the case of a one-dimensional model for the point process of incoming market orders of the Bund futures. Errais et al. (2010) underlines the  
305 computational tractability of an exponential kernel in the context of portfolio credit risk. In the case of earthquakes, which is one of the original applications of Hawkes processes, the goodness-of-fit of power-law kernel has been enlighten by Ogata (1988).

310 In the context of criminality (analysis of a dataset providing civilian deaths due to insurgent activity in Iraq), Lewis & Mohler (2011) model the dates of

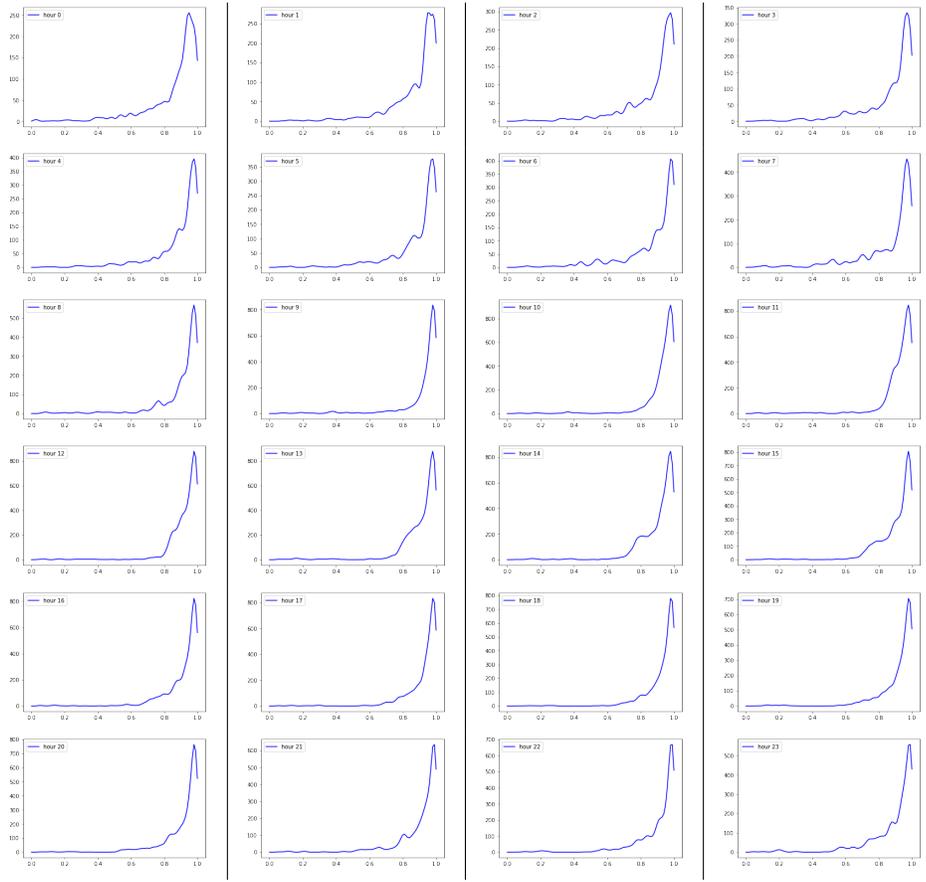


Figure 7: Non-parametric detection of a change-point (kernel estimator  $\hat{h}$ )

crimes with a Hawkes process, and use a nonparametric EM algorithm to estimate the baseline intensity  $\mu$  and the kernel  $\varphi$ . They assume the timescale over which  $\mu$  is changing is larger than the timescale over which  $\varphi$  decays. For the electricity market, Becker et al. (2013) uses an exponential kernel with few discussion about this choice (except that is very tractable for the computations).

Several strategies were investigated to perform a non-parametric estimation of the kernel assuming that the point process is stationary and observed on a long time range. Reynaud-Bouret et al. (2010) uses a penalized contrast to derive an adaptive estimation of the kernel using piecewise constant functions. (However, in practice, computations involved in the implementation are quite costly in time.) In Kirchner (2016), a stationary Hawkes process is approximated in distribution by a INAR( $\infty$ ) process. Then, in the companion paper (Kirchner (2017)), a nonparametric estimation of  $\varphi$  is derived by the estimation of the coefficients of an autoregressive process. Finally, Bacry et al. (2012) use deconvolution methods to estimate the kernel, and Bacry & Muzy (2016) combine them with kernel estimation in the case of a marked point process.

All of these strategies have to be reconsidered due to another asymptotic framework, because dates of trades, assumed to be observations of a point process, are observed on a time interval of fixed length.

The non-parametric procedure, presented in this section, is devoted to the estimation of the kernel, and derived from Bai et al. (2015) with a slight modification to take into account the piecewise constant baseline. In this section, we assume that we have preliminary estimated  $\delta$ . The idea is to write the log-likelihood

$$\begin{aligned} \ell_n(\mu_0, \mu_1, \varphi) = & \sum_{k=1}^n \left\{ \left( \sum_{t_i^k < \delta} \ln(\mu_0 + \sum_{t_j^k < t_i^k} \varphi(t_i^k - t_j^k)) - \delta \mu_0 \right) \right. \\ & + \left( \sum_{t_i^k \geq \delta} \ln(\mu_1 + \sum_{t_j^k < t_i^k} \varphi(t_i^k - t_j^k)) - (1 - \delta) \mu_1 \right) \\ & \left. - \sum_{t_i^k} \int_0^{1-t_i^k} \varphi(u) du \right\} \end{aligned}$$

and built an estimator  $(\hat{\mu}_0, \hat{\mu}_1, \hat{\varphi})$  by maximizing it, assuming that  $\hat{\varphi}$  is a B-spline approximation of the function  $\varphi$ . Hence, the problem is reduced to a

finite-dimensional optimization problem for the computation of the estimator (a B-spline is a linear combination of basis functions, so only the vector of coefficients has to be estimated).

A B-spline basis is characterized by the regularity  $r$  of the functions,  $d \geq r+1$  the order of the spline basis functions and  $\kappa_n$  (depending on the sample size) the size of the basis. A knot sequence  $\xi$  is a sequence of length  $\kappa_n + d$  such that

$$0 = \xi_1 = \dots = \xi_d < \xi_{d+1} < \dots < \xi_{\kappa_n+1} = \dots = \xi_{\kappa_n+d} = 1$$

and in practice, we will assume that the internal knots (distinct values of  $\xi$ ) are regularly spaced. Let  $B(t) = (B_1(t), \dots, B_{\kappa_n}(t))'$  the B-spline basis functions associated with this knot sequence.

**Definition 3.** *The set of B-spline basis functions  $B_i^d(t)$  of order  $d$  ( $1 \leq i \leq \kappa_n$ ) is defined recursively by*

$$B_i^k(t) = \frac{t - \xi_i}{\xi_{i+k-1} - \xi_i} B_i^{k-1}(t) + \frac{\xi_{i+k} - t}{\xi_{i+k} - \xi_{i+1}} B_{i+1}^{k-1}(t)$$

and the initial condition  $B_i^1(t) = \mathbf{1}_{[\xi_i, \xi_{i+1})}(t)$ .

One important result about B-splines is that, for  $\gamma \in \mathbb{R}^{\kappa_n}$  such that  $\gamma_1 \geq \dots \geq \gamma_{\kappa_n} \geq 0$  (positive decreasing sequence), the function  $S$  defined by  $S(t) = \gamma' B(t)$  is a non-negative and non-increasing function over  $[0, 1]$ . Then, a B-spline estimator of the kernel has these properties.

**Remark 5.** *One of the advantage of the B-spline method is to simplify the computation of the integrals involved in the log-likelihood by an evaluation of a linear combination of B-spline functions of order  $d + 1$ . Hence, the evaluation of the log-likelihood is very fast once the basis functions are computed.*

Figures 8 and 9 present the results of the semi-parametric estimation of  $(\mu_0, \mu_1, \varphi)$ , using B-splines of order  $d = 3$  and 16 internal nodes in  $[0, 1]$  (a grid of stepsize  $\frac{1}{16}$ ) o, simulations. In both cases, the values of  $\mu_0$  and  $\mu_1$  are

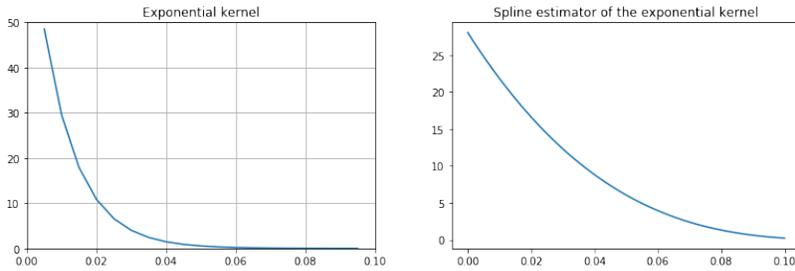


Figure 8: Exponential kernel and its non-parametric spline estimation (simulation)

correctly estimated (and a theoretical result for the rate of convergence can be found in Bai et al. (2015)). The optimization of the log-likelihood is performed with the Python library `scipy.optimize` and the computation of the B-splines is based on the Python function `scipy.interpolate.BSpline`.

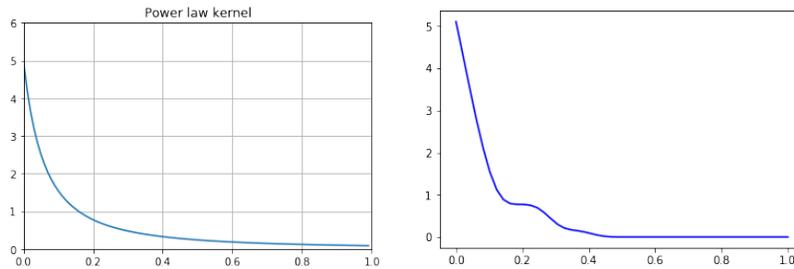


Figure 9: Power-law kernel and its non-parametric spline estimation (simulation)

### 355 5.1. Estimation on different hours

In a second step of the study started in Section 4.2, a non-parametric estimation of the kernel is performed for each hour using the B-spline method. As we don't have a precise estimation of  $\delta_0$  at this time, and because the value of  $\delta$  chosen for the baseline estimation does not give a real impact on the kernel  
 360 estimation, we let  $\delta = 0.85$  for sake of simplicity. Figure 10 shows the results. For each hour, the shape of the estimated kernel is a fast decreasing function, which vanishes during the second part of the period. For this reason, we prefer to deal with an exponential kernel in the sequel.

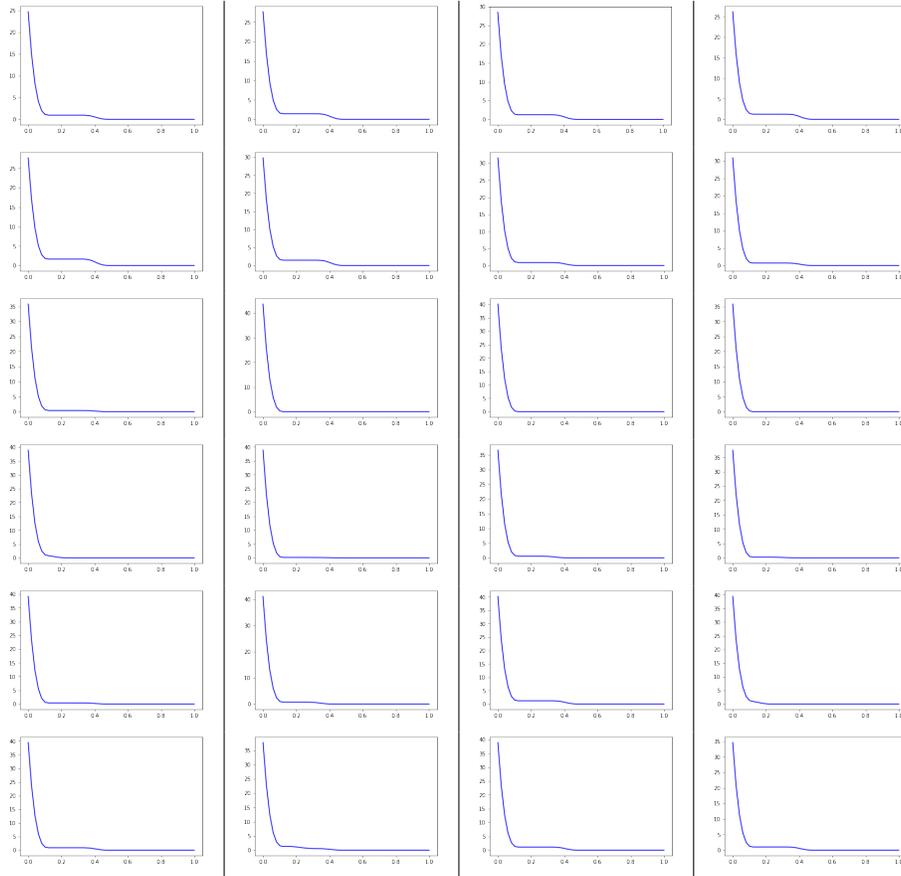


Figure 10: Non-parametric estimation of the kernel (B-splines,  $\delta_0 = 0.85$ )

## 6. Parametric estimation and comparison of the results

365 The aim of this section is to compare the results obtained by a non-parametric  
method to those obtained for the maximum-likelihood estimator. Due to the  
detection of a change-point in the baseline function, the question of a likelihood  
ratio test is addressed at a first look. The discussion is about two different rates  
of convergence : a  $\sqrt{n}$  rate for the parameters for which the log-likelihood is  
370 twice differentiable, and a  $n$  rate for the estimation of  $\delta$ . Numerical results are  
provided, based on a `Python` library.

### 6.1. Likelihood ratio for testing and estimation

In a recent article (Chernoyarov et al. (2018)), a problem of change-point  
detection for inhomogeneous Poisson processes is presented. We follow the lines  
375 of this article to study the asymptotic properties of the change-point test in the  
case of a Hawkes process observed on  $[0, 1]$ . The likelihood ratio is of major  
interest, both for the the asymptotic properties of the MLE, and for testing if  
a change-point occurs.

The parameter  $\delta$  is supposed to be unknown, but other parameters (involved  
in  $\mu$  and  $\varphi$ ) are assumed to be known. We discuss later about a joint estima-  
tion on numerical studies. For one observation  $N_i$  of the Hawkes process, the  
intensity is

$$\lambda_i(t) = \mu_0 + \int_0^t \varphi(t-s) dN_i(s) + (\mu_1 - \mu_0) \mathbf{1}_{(t \geq \delta)} = \lambda_i^0(t) + (\mu_1 - \mu_0) \mathbf{1}_{(t \geq \delta)}$$

Let  $\ln \left( \frac{d\mathbf{P}_\delta}{d\mathbf{P}_{\delta_1}}(N_i) \right)$  the logarithm of the likelihood ratio between the distribution  
380 of one process with a change-point at time  $\delta$  and one with a change-point at  
time  $\delta_1$ . With  $\theta_1 = \mu_1 - \mu_0$ ,

$$\begin{aligned} \ln \left( \frac{d\mathbf{P}_\delta}{d\mathbf{P}_{\delta_1}}(N_i) \right) &= \int_0^1 \ln \left( \frac{\lambda_i^0(t) + \theta_1 \mathbf{1}_{(t \geq \delta)}}{\lambda_i^0(t) + \theta_1 \mathbf{1}_{(t \geq \delta_1)}} \right) dN_i(t) - \int_0^1 \theta_1 (\mathbf{1}_{(t \geq \delta)} - \mathbf{1}_{(t \geq \delta_1)}) dt \\ &= \int_0^1 \ln \left( \frac{\lambda_i^0(t) + \theta_1 \mathbf{1}_{(t \geq \delta)}}{\lambda_i^0(t) + \theta_1 \mathbf{1}_{(t \geq \delta_1)}} \right) dN_i(t) - \theta_1 (\delta_1 - \delta) \end{aligned}$$

This is slightly different from the log-likelihood of Equation 2, where the likelihood is written with respect to a homogeneous Poisson process of intensity 1. Then, with  $n$  i.i.d. observations, the test statistics becomes

$$\mathcal{L}(\delta) = \sum_{i=1}^n \ln \left( \frac{d\mathbf{P}_\delta}{d\mathbf{P}_1}(N_i) \right)$$

(with  $d\mathbf{P}_1$  the distribution of a process with constant baseline, but any arbitrary value can be chosen). Let  $\hat{\delta}$  the MLE defined by

$$\hat{\delta} \in \arg \max_{\delta \in (0,1)} \mathcal{L}(\delta)$$

From a numerical point of view, and because  $\mathcal{L}$  is not a regular function, the maximization is performed over a grid. Therefore, testing if the baseline is constant, or not, can be decided from the values of this ratio : under the null hypothesis  $\mathcal{H}_0$ ,  $\mathcal{L}(\delta) \leq 0$  for  $\delta \in [0, 1]$ , whereas under  $\mathcal{H}_1(\delta_0)$ ,  $\mathcal{L}(\delta)$  reaches its maximum for  $\delta = \delta_0$ . Figure 11 illustrates this fact with simulated data (for these simulations, an exponential kernel has been chosen).

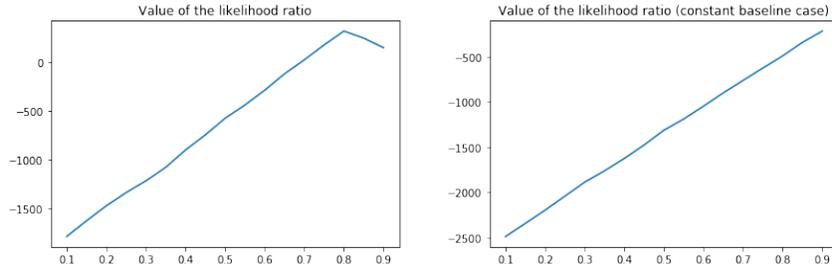


Figure 11: Values of  $\mathcal{L}(\delta)$  (left : change-point at 0.8, right : constant baseline)

Moreover, in the literature of change-point estimation, asymptotic properties of the estimators are often derived from the asymptotic behaviour of the normalized likelihood ratio (Ibragimov & Has'minskii (1981) , Chernoyarov et al. (2018)). In particular, the rate of convergence of the change-point estimator is faster than the other parameters when the likelihood is not smooth. To be pre-

cise, if  $\frac{dP(\delta_0 + \frac{u}{n})}{dP(\delta_0)}$  has a limit in distribution as  $n \rightarrow \infty$ , then  $n(\hat{\delta}_n - \delta)$  converges in distribution, with  $\hat{\delta}$  the MLE estimator of  $\delta$ . To assess the limit for a self-  
 395 exciting process, we performed simulations (100 replications of 30 independent trajectories with  $\delta_0 = 0.8$  and an exponential kernel) to draw an histogram of the empirical distribution of  $n(\hat{\delta} - \delta_0)$  and confirm empirically the rate of convergence. Figure 12 shows the histogram of the limit distribution. Unfortunately, the complete proof is beyond our reaching.

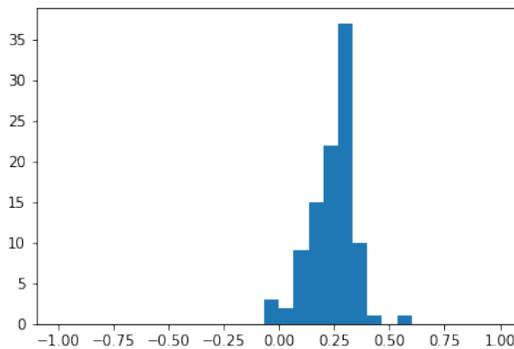


Figure 12: Empirical distribution of  $n(\hat{\delta}_n - \delta)$  (simulated data)

400 *6.2. Numerical results*

In order to perform the numerical maximization of the log-likelihood including a change-point detection,

- for fixed values of  $\delta \in \{j\Delta, j = 1, \dots, N\}$ , the MLE estimation of other parameters is done by a classical algorithm (method 'L-BFGS-B' from  
 405 `scipy.optimize.minimize`),
- these estimated values are stored, with the value of the log-likelihood,
- $\hat{\delta} \in \arg \max \mathcal{L}(\delta)$  is computed, with the associated values of the other parameters.

. On simulated datasets, this method gives positive results. In Figure 13, the  
 410 data is simulated with

- an exponential kernel ( $\beta_0 = 80$  and  $\beta_1 = 100$ )
- a changepoint in the baseline at  $\delta_0 = 0.8$  ( $\mu_0 = 5$  and  $\mu_1 = 50$ )

The estimated values are :

$$\hat{\mu}_0 = 4.62 \quad , \quad \hat{\mu}_1 = 45.09 \quad , \quad \hat{\beta}_0 = 81.20 \quad , \quad \hat{\beta}_1 = 100.37$$

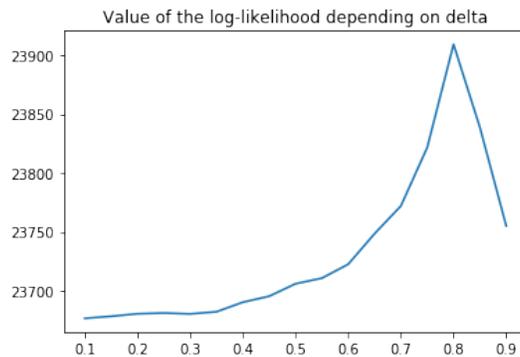


Figure 13: Maximization of the log-likelihood with respect to  $\delta$  (true value = 0.8)

In the case of data simulated with constant baseline ( $\mu_0 = 5$ ), the estimated  
 values are

$$\hat{\mu}_0 = 5.21 \quad , \quad \hat{\beta}_0 = 80.21 \quad , \quad \hat{\beta}_1 = 97.24$$

and the values of the log-likelihood belong to  $[10196, 10200]$ . Hence, it is easy  
 to distinguish if there is a change point or not (*i.e.* if there is a clear spike in  
 415  $\mathcal{L}(\delta)$  or not).

Figure 15 shows that in the case of an exponential kernel, the opposite of the  
 log-likelihood (which is provided to the minimizer of `Scipy`) is convex around its  
 minimum. Hence, the computation of the MLE is well performed. However, in  
 the case of a power-law kernel, the opposite of the log-likelihood is flat around  
 420 its minimum, as shown in Figure 16. Then, the computation of the MLE is

harder, and the numerical errors of approximation lead to important errors for the MLE.

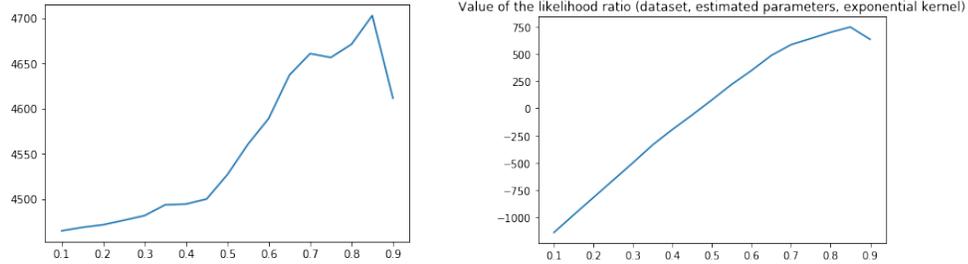


Figure 14: Left : Maximization of the log-likelihood with respect to  $\delta$ . Right : Likelihood ratio. (exponential kernel, April 2015)

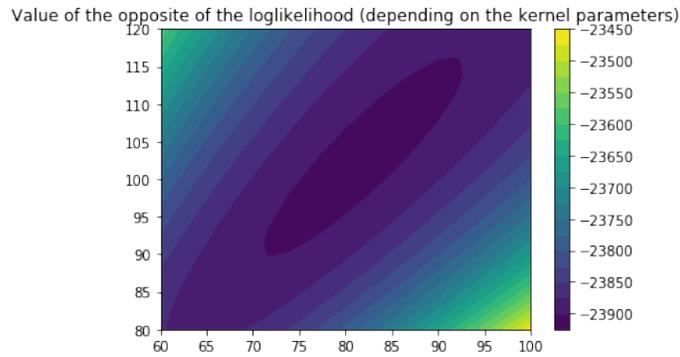


Figure 15: Influence of the values of the kernel parameters on the log-likelihood (exponential kernel)

### 6.3. Estimation on different hours

Figure 17 shows for each hour the value of the log-likelihood, as a function of  $\delta$ , where the other parameters are estimated by a classical method of maximization. Pictures show a clear maximum on the grid, and we have chosen a thinner grid between 0.7 and 0.9. Due to the previous studies on simulated datasets, MLE estimators for the parameters of a model based on an exponential kernel are given in Table 2. (Warning : as the length of the trading period

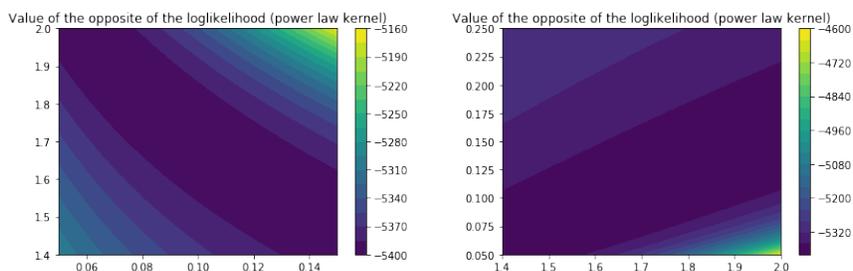


Figure 16: Influence of the values of the kernel parameters on the log-likelihood (power law kernel)

430 increases from the first to the last hour, the estimated values of  $\alpha$  and  $\beta$  cannot be compared between different hours). However, the ratio  $\alpha/\beta$  is equal to  $\|\varphi\|_1$  in any case.

Asymptotic confidence intervals in Table 2 are based on an approximation of the Fisher information matrix (see Theorem 1) and the multivariate central  
 435 limit theorem, assuming that we can replace  $\delta_0$  by  $\hat{\delta}$  (due to a faster rate of convergence) and apply the theorem for regular cases.

These confidence intervals reported in Table 2 show that the parameters of the baseline  $(\mu_0, \mu_1)$  and the decay parameter  $(\beta)$  are estimated with a half-length of the interval which is about 10% of the value of the estimator. Hence  
 440 these estimations (with 30 observed trajectories) are not of good precision. Conversely, the estimation of  $\alpha$  is very good, and the ratio  $\alpha/\beta$  is quite stable.

## 7. Conclusion

As a self-excitement phenomenon has been enlighten in the intraday electricity market, the question of the understanding of its causes is now a perspective.  
 445 It is plausible to consider economic reasons (convergence of prices in the order book at the end of the trading period due to an adjustment of supply and demand). Technical implications may also be considered, as electricity production requires a fine adjustment : a better forecasting of the weather or the supply capacities of renewable producers may be obtained at the end of the period.

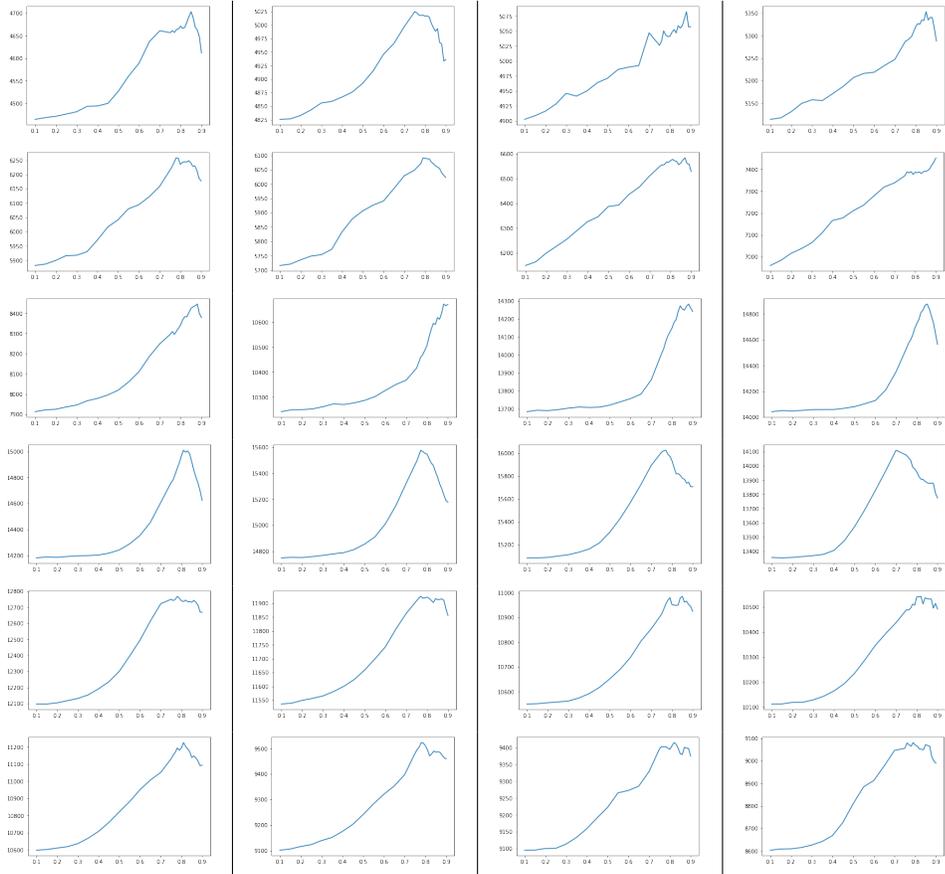


Figure 17: Maximum Likelihood Estimation of  $\delta$  (value of  $\mathcal{L}(\delta)$ )

Hour	$\hat{\mu}_0 (\pm \frac{1.96\sigma}{\sqrt{n}})$	$\hat{\mu}_1 (\pm \frac{1.96\sigma}{\sqrt{n}})$	$\hat{\alpha} (\pm \frac{1.96\sigma}{\sqrt{n}})$	$\hat{\beta} (\pm \frac{1.96\sigma}{\sqrt{n}})$	$\hat{\alpha}/\hat{\beta}$	$\hat{\delta}$
0	9.24 (1.20)	120.58 (10.98)	1042.93 (1.37)	3046.76 (400.02)	0.34	0.85
1	6.21 (1.04)	86.29 (7.19)	1092.30 (0.84)	2955.33 (346.44)	0.37	0.75
2	12.10 (1.34)	149.64 (13.60)	1372.76 (1.82)	4282.14 (560.11)	0.32	0.88
3	10.75 (1.28)	135.79 (11.61)	1456.00 (1.59)	4482.46 (589.64)	0.32	0.85
4	7.65 (1.13)	109.80 (8.57)	1693.60 (0.97)	4481.37 (492.68)	0.38	0.78
5	8.16 (1.16)	109.20 (8.61)	2116.55 (1.04)	5646.01 (614.23)	0.37	0.79
6	12.01 (1.34)	148.84 (12.92)	2238.90 (1.47)	5587.76 (579.94)	0.40	0.87
7	14.09 (1.43)	196.01 (17.15)	2176.80 (1.63)	5359.41 (579.66)	0.41	0.90
8	8.40 (1.12)	204.08 (16.19)	2423.27 (0.92)	5492.96 (517.73)	0.44	0.88
9	7.04 (1.02)	238.33 (18.42)	2194.51 (0.72)	4470.30 (422.32)	0.49	0.88
10	8.13 (1.10)	336.96 (21.31)	2996.66 (0.74)	6641.72 (546.85)	0.45	0.88
11	5.95 (0.95)	300.22 (17.61)	3440.93 (0.54)	7787.35 (625.70)	0.44	0.85
12	3.98 (0.80)	241.54 (14.05)	3309.39 (0.35)	7204.63 (560.89)	0.46	0.81
13	4.13 (0.84)	216.32 (12.03)	3405.75 (0.35)	7637.79 (573.39)	0.46	0.77
14	6.17 (1.02)	221.68 (12.05)	3582.57 (0.52)	8236.99 (608.84)	0.43	0.77
15	3.17 (0.76)	160.73 (8.94)	3219.25 (0.28)	7369.48 (576.16)	0.44	0.70
16	6.02 (1.00)	177.16 (10.98)	3425.26 (0.53)	7477.16 (576.09)	0.46	0.78
17	4.09 (0.83)	142.32 (10.09)	2066.99 (0.35)	3900.27 (294.16)	0.53	0.77
18	7.80 (1.09)	229.72 (15.35)	2900.38 (0.78)	6875.13 (599.48)	0.42	0.85
19	7.42 (1.09)	184.69 (12.50)	2525.62 (0.72)	5799.84 (488.55)	0.44	0.82
20	7.08 (1.07)	186.55 (12.19)	3018.46 (0.65)	6913.67 (541.54)	0.44	0.81
21	5.45 (0.95)	141.72 (10.06)	2012.92 (0.54)	4443.72 (380.92)	0.45	0.78
22	9.37 (1.22)	187.69 (12.24)	3910.28 (1.03)	10881.17 (1032.8)	0.36	0.82
23	9.24 (1.23)	153.88 (10.26)	3405.85 (0.98)	9086.60 (897.83)	0.37	0.79

Table 2: Estimated values of the parameters for various hours of the day (exponential kernel)

450 Finally, this work focused on the dates of trades, especially with self-excitation  
as a source of endogeneity, and the last perspective would be to analyze the price  
process on the European market, taking into account exogenous variables and  
seasonality, as in Becker et al. (2013).

In conclusion, I would like to thank Prof. Rüdiger Kiesel and his team for  
455 his welcoming during my visit in Essen, and the discussions we had about the  
electricity market.

## References

- Bacry, E., Dayri, K., & Muzy, J.-F. (2012). Non-parametric kernel estimation  
for symmetric Hawkes processes. application to high frequency financial data.  
460 *The European Physical Journal B*, *85*, 157.
- Bacry, E., Delattre, S., Hoffmann, M., & Muzy, J.-F. (2013a). Modelling mi-  
crostructure noise with mutually exciting point processes. *Quantitative Fi-  
nance*, *13*, 65–77.
- Bacry, E., Delattre, S., Hoffmann, M., & Muzy, J.-F. (2013b). Some limit the-  
465 orems for Hawkes processes and application to financial statistics. *Stochastic  
Processes and their Applications*, *123*, 2475–2499.
- Bacry, E., Mastromatteo, I., & Muzy, J.-F. (2015). Hawkes processes in finance.  
*Market Microstructure and Liquidity*, *1*, 1550005.
- Bacry, E., & Muzy, J.-F. (2014). Hawkes model for price and trades high-  
470 frequency dynamics. *Quantitative Finance*, *14*, 1147–1166.
- Bacry, E., & Muzy, J.-F. (2016). First- and second-order statistics characteriza-  
tion of Hawkes processes and non-parametric estimation. *IEEE Transactions  
on Information Theory*, *62*, 2184–2202.
- Bai, F., Chen, F., & Chen, K. (2015). Semiparametric estimation of a self-  
475 exciting regression model with an application in recurrent event data analysis.  
*Statistica Sinica*, (pp. 1503–1526).

- Becker, R., Clements, A. E., & Zainudin, W. (2013). Modeling electricity price events as point processes. *Journal of Energy Markets*, *6*, 99–140.
- Chen, F., & Hall, P. (2013). Inference for a nonstationary self-exciting point  
 480 process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability*, *50*, 1006–1024.
- Chen, F., & Hall, P. (2016). Nonparametric estimation for self-exciting point processes - a parsimonious approach. *Journal of Computational and Graphical Statistics*, *25*, 209–224.
- 485 Chernoyarov, O., Kutoyants, Y. A., & Top, A. (2018). On multiple change-point estimation for poisson process. *Communications in Statistics-Theory and Methods*, *47*, 1215–1233.
- Da Fonseca, J., & Zaatour, R. (2014). Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*,  
 490 *34*, 548–579.
- Da Fonseca, J., & Zaatour, R. (2015). Clustering and mean reversion in a hawkes microstructure model. *Journal of Futures Markets*, *35*, 813–838.
- Errais, E., Giesecke, K., & Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, *1*, 642–665.
- 495 Ibragimov, I., & Has'minskii, R. (1981). Z.(1981). statistical estimation, asymptotic theory. *Springer, New York*, *1*, I4.
- Kiesel, R., & Paraschiv, F. (2017). Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, *64*, 77–90.
- Kirchner, M. (2016). Hawkes and inar ( $\infty$ ) processes. *Stochastic Processes and their Applications*, *126*, 2494–2525.  
 500
- Kirchner, M. (2017). An estimation procedure for the hawkes process. *Quantitative Finance*, *17*, 571–595.

- Lewis, E., & Mohler, G. (2011). A nonparametric em algorithm for multiscale hawkes processes. *Journal of Nonparametric Statistics*, 1, 1–20.
- 505 Lewis, E., Mohler, G., Brantingham, P. J., & Bertozzi, A. L. (2012). Self-exciting point process models of civilian deaths in iraq. *Security Journal*, 25, 244–264.
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*,  
 510 83, 9–27.
- Rambaldi, M., Filimonov, V., & Lillo, F. (2018). Detection of intensity bursts using hawkes processes: An application to high-frequency financial data. *Physical Review E*, 97, 032318.
- Reynaud-Bouret, P., Schbath, S. et al. (2010). Adaptive estimation for hawkes  
 515 processes; application to genome analysis. *The Annals of Statistics*, 38, 2781–2822.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* volume 3. Cambridge university press.

## Appendix

520 In this section, the proofs of some results are gathered to ease the reading of the report.

### *Appendix.1. Asymptotic properties of the sum-process*

This section provides asymptotic properties of the sum-process. The most important idea to keep in mind is that two types of convergence occur, at different rates :

525

1. A law-of-large-numbers result for  $n^{-1}\mathcal{N}_n$ , with a deterministic limit ;
2. A central-limit-theorem result for  $n^{-1/2}\mathcal{M}_n$ , with a Gaussian limit.

These computations are useful to derive the positivity of the covariance matrix of the observations, and assert a model with self-excitement.

530 . First, a uniform law of large numbers is established for  $n^{-1}\mathcal{N}_n$ .

**Proposition 7.** For  $t \in [0, 1]$

$$\frac{\mathcal{N}_n(t)}{n} \xrightarrow{P} \mathbf{E}(N(t)) = H(t)$$

Moreover, the convergence is uniform in  $t$  :

$$\sup_{t \in [0, 1]} \left| \frac{\mathcal{N}_n(t)}{n} - H(t) \right| \xrightarrow{P} 0$$

The following proposition is the key of further computations for the convergence in distribution of the martingale part.

**Proposition 8.** Let  $f$  and  $g$  two continuous functions defined on  $[0, 1]$ , and denote by  $F$  and  $G$  their antiderivatives. For  $(s, t) \in [0, 1]^2$ ,

$$\mathbf{E} \left( \int_0^s f(s-u)M(u)du \right) = 0$$

and

$$C(f(s-\cdot), g(t-\cdot)) = \mathbf{E} \left( \int_0^s f(s-u)M(u)du \right) \left( \int_0^t g(t-u)M(u)du \right) = \int_0^{s \wedge t} F(s-u)G(t-u)h(u)du$$

In particular,

$$\mathbf{V}(M(t) + \psi \star M(t)) = (\mathbf{1} + \Psi)^2 \star h(t) = \int_0^t (\mathbf{1}_{[0, t]}(t-u) + \Psi(t-u))^2 h(u)du$$

where  $\Psi$  is an antiderivative of  $\psi$ .

535 . The limit in distribution of the martingale part of  $\mathcal{N}$  can be determined using Proposition 8, and a Central Limit Theorem.

**Theorem 2.** For  $(s, t) \in [0, 1]^2$ ,

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \mathcal{M}_n(s) \\ \psi \star \mathcal{M}_n(s) \\ \mathcal{M}_n(t) \\ \psi \star \mathcal{M}_n(t) \end{pmatrix} \xrightarrow{d} \mathcal{N} \left( \mathbf{0}; \begin{pmatrix} \mathbf{1} \star h(s) & \Psi \star h(s) & \mathbf{1} \star h(s \wedge t) & (a) \\ \Psi \star h(s) & \Psi^2 \star h(s) & (b) & (c) \\ \mathbf{1} \star h(s \wedge t) & (b) & \mathbf{1} \star h(t) & \Psi \star h(t) \\ (a) & (c) & \Psi \star h(t) & \Psi^2 \star h(t) \end{pmatrix} \right)$$

with

$$\begin{aligned} (a) &= \int_0^{s \wedge t} \Psi(t-u)h(u)du \\ (b) &= \int_0^{s \wedge t} \Psi(s-u)h(u)du \\ (c) &= \int_0^{s \wedge t} \Psi(s-u)\Psi(t-u)h(u)du \end{aligned}$$

**Corollary 1.** For  $(s, t) \in [0, 1]^2$ ,

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \mathcal{M}_n(s) + \psi \star \mathcal{M}_n(s) \\ \mathcal{M}_n(t) + \psi \star \mathcal{M}_n(t) \end{pmatrix} \xrightarrow{d} \mathcal{N} \left( \mathbf{0}; \begin{pmatrix} \mathbf{c}(s, s) & \mathbf{c}(s, t) \\ \mathbf{c}(t, s) & \mathbf{c}(t, t) \end{pmatrix} \right)$$

with

$$\mathbf{c}(s, t) = \int_0^{s \wedge t} (1 + \Psi(s-u))(1 + \Psi(t-u))h(u)du$$

From now on, the properties of the limit distribution are investigated. The idea is to consider a Gaussian process with the same marginal distributions.

**Definition 4.** Let  $(B_t), t \in [0, 1]$  a standard Brownian motion on  $[0, 1]$ , and

$$Y_t = \int_0^t \sqrt{h(u)}dB_u \quad (.1)$$

**Lemma 5.** We have :

- 540
- $d\langle Y \rangle_t = h(t)dt$
  - $Y_t \sim \mathcal{N}(0, \int_0^t h(u)du)$

- for  $(s, t) \in [0, 1]^2$

$$\text{Cov}(Y_s, Y_t) = \int_0^{s \wedge t} h(u) du$$

Hence, the diffusion process  $(Y_t)$  has the same properties as the limiting distribution of  $n^{-1/2}\mathcal{M}_n(t)$ .

**Proposition 9.** For  $t \in [0, 1]$ , let

$$Z_t = Y_t + (\psi \star Y)(t) = Y_t + \int_0^t \psi(t-u) Y_u du$$

Then  $(Z_t)$  is a centered Gaussian process with covariance  $\mathbf{c}(s, t)$ . Furthermore,

$$\forall t \in [0, 1], \quad Z_t = Y_t + \int_0^t \varphi(t-u) Z_u du$$

These results imply :

- 545 •  $n^{-1}\mathcal{N}_n(t)$  converges to  $H(t) = m(t) + \int_0^t \varphi(t-s)H(s)ds$  in  $L^2$
- $n^{-1/2}(\mathcal{M}_n(t) + \psi \star \mathcal{M}_n(t))$  converges to  $Z_t = Y_t + \int_0^t \varphi(t-s)Z_s ds$  in distribution

and we can draw the analogy between  $(M_t)$  and  $(Y_t)$ . This is strongly linked with Proposition 5 (Auto-Regressive projection) of Bacry et al. (2015).

550 *Proof of Proposition 4.* Let  $N_1, \dots, N_n$   $n$  independent Hawkes processes on  $[0, 1]$ , of respective intensities  $\lambda_1, \dots, \lambda_n$ . Let  $\mathcal{N}(t) = N_1(t) + \dots + N_n(t)$ .

For  $h > 0$ , using independent events,

$$\begin{aligned} \mathbf{P}(\mathcal{N}(t+h) - \mathcal{N}(t) = 0 | \mathcal{F}_t) &= \mathbf{P}\left(\bigcap_{i=1}^n \{N_i(t+h) - N_i(t) = 0\} | \mathcal{F}_t\right) \\ &= \prod_{i=1}^n \mathbf{P}(N_i(t+h) - N_i(t) = 0 | \mathcal{F}_t) \\ &= \prod_{i=1}^n (1 - \lambda_i(t)h + o(h)) \\ &= 1 - \left(\sum_{i=1}^n \lambda_i(t)\right)h + o(h) \end{aligned}$$

and, using disjunct events,

$$\begin{aligned}
\mathbf{P}(\mathcal{N}(t+h) - \mathcal{N}(t) = 1 | \mathcal{F}_t) &= \mathbf{P}\left(\bigcup_{i=1}^n \{N_i(t+h) - N_i(t) = 1\} \cap \bigcap_{j \neq i} \{N_j(t+h) - N_j(t) = 0\} | \mathcal{F}_t\right) \\
&= \sum_{i=1}^n \mathbf{P}\left(\{N_i(t+h) - N_i(t) = 1\} \cap \bigcap_{j \neq i} \{N_j(t+h) - N_j(t) = 0\} | \mathcal{F}_t\right) \\
&= \sum_{i=1}^n (\lambda_i(t)h + o(h)) \prod_{j \neq i} (1 - \lambda_j(t)h + o(h)) \\
&= \sum_{i=1}^n \lambda_i(t)h + o(h)
\end{aligned}$$

Hence,  $\mathcal{N}$  is a Hawkes process with intensity  $\Lambda = \sum_{i=1}^n \lambda_i$ .  $\square$

*Proof of Lemma 2..* Let  $t \in [0, 1]$ . As

$$H(t) = \int_0^t \mu(s) ds + \int_0^t \psi(s) m(t-s) ds$$

we have (by interchanging the integral and the derivative)

$$H'(t) = \mu(t) + \psi(t)m(0) + \int_0^t \psi(s)m'(t-s) ds = h(t)$$

555 since  $m(0) = 0$ .  $\square$

*Proof of Proposition 7..* For  $t \in [0, 1]$ ,

$$\frac{1}{n} \mathcal{N}(t) = \frac{1}{n} \sum_{i=1}^n N_i(t)$$

and by the law of large numbers, as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{i=1}^n N_i(t) \xrightarrow{L^2} \mathbf{E}(N(t))$$

and the integral formula for  $\mathbf{E}(N(t))$  is given by Lemma 1. For the uniform convergence, see for example Chen & Hall (2016).  $\square$

*Proof of Proposition 8..* With integrations by parts,

$$C(f(s-), g(t-)) = \mathbf{E} \left( \int_0^1 \mathbf{1}_{[0,s]}(s-u)F(s-u)dM(u) \right) \left( \int_0^1 \mathbf{1}_{[0,t]}(t-u)G(t-u)dM(u) \right)$$

Using the Itô's isometry and  $d[M](u) = dN(u)$ ,

$$\begin{aligned} C(f(s-), g(t-)) &= \mathbf{E} \left( \int_0^1 \mathbf{1}_{[0,s]}(s-u)\mathbf{1}_{[0,t]}(t-u)F(s-u)G(t-u)dN(u) \right) \\ &= \int_0^{s \wedge t} F(s-u)G(t-u)h(u)du \end{aligned}$$

□

560 *Proof of Theorem 2..* Based on the multivariate CLT and the result of Proposition 8.

□

*Proof of Proposition 7..* Using the martingale decomposition  $dN(s) = dM(s) + \lambda(s)ds$ , we have :

$$\begin{aligned} \int_{t_1}^{t_2} \int_{t_3}^{t_4} (dM(s) + \lambda(s)ds)(dM(u) + \lambda(u)du) &= \int_{t_1}^{t_2} \int_{t_3}^{t_4} dM(s)dM(u) + \int_{t_1}^{t_2} dM(s) \int_{t_3}^{t_4} \lambda(u)du \\ &\quad + \int_{t_1}^{t_2} \lambda(s)ds \int_{t_3}^{t_4} dM(u) + \int_{t_1}^{t_2} \int_{t_3}^{t_4} \lambda(s)\lambda(u)dsdu \end{aligned}$$

By Itô's isometry,

$$\mathbf{E} \left( \int_{t_1}^{t_2} dM(s) \int_{t_3}^{t_4} dM(u) \right) = \int_{t_1 \vee t_3}^{t_2 \wedge t_4} h(s)ds$$

565 As  $\lambda(t) = h(t) + \psi \star dM(t)$ ,

$$\begin{aligned} \mathbf{E} \left( \int_{t_1}^{t_2} dM(s) \int_{t_3}^{t_4} \lambda(u)du \right) &= \mathbf{E} \left( \int_{t_1}^{t_2} dM(s) \int_{t_3}^{t_4} \psi \star dM(u)du \right) \\ &= \int_{t_1}^{t_2} \int_{t_3}^{t_4} \psi(u-s)h(s)dsdu \end{aligned}$$

using Fubini's theorem and Itô's isometry.

Finally,

$$\begin{aligned} \mathbf{E} \left( \int_{t_1}^{t_2} \int_{t_3}^{t_4} \lambda(s)\lambda(u)dsdu \right) &= \int_{t_1}^{t_2} \int_{t_3}^{t_4} h(s)h(u)dsdu \\ &+ \int_{t_1}^{t_2} \int_{t_3}^{t_4} \underbrace{\mathbf{E} \left( \int_0^s \psi(s-v)dM(v) \int_0^u \psi(u-w)dM(w) \right)}_{=\gamma(s,u)} dsdu \end{aligned}$$

and the conclusion comes from :

$$\gamma(s, u) = \int_{s \wedge u} \psi(s-v)\psi(u-v)h(v)dv.$$

□

*Proof of Theorem 1..* We consider the normalized score function

$$\frac{1}{n}S_n(\theta) = \frac{1}{n} \sum_{k=1}^n \int_0^1 \frac{\partial_\theta \lambda_k(t; \theta)}{\lambda_k(t; \theta)} dM_k(t; \theta)$$

By the law of large numbers,

$$\frac{1}{n}S_n(\theta) \xrightarrow{\mathbf{P}} s(\theta, \theta_0) = \mathbf{E}_{\theta_0} \left( \int_0^1 \frac{\partial_\theta \lambda(t; \theta)}{\lambda(t; \theta)} (dN(t) - \lambda(t; \theta)dt) \right)$$

and

$$s(\theta, \theta_0) = \mathbf{E}_{\theta_0} \left( \int_0^1 \frac{\partial_\theta \lambda(t; \theta)}{\lambda(t; \theta)} (\lambda(t; \theta_0) - \lambda(t; \theta))dt \right).$$

In addition,

$$\partial_\theta s(\theta, \theta_0) = \mathbf{E}_{\theta_0} \left( \int_0^1 \left( \frac{\partial_{\theta\theta'}^2 \lambda(t; \theta)}{\lambda(t; \theta)} - \frac{\partial_\theta \lambda(t; \theta)^{\otimes 2}}{\lambda(t; \theta)^2} \right) \lambda(t; \theta_0)dt \right) - \mathbf{E}_{\theta_0} \left( \int_0^1 \partial_{\theta\theta'}^2 \lambda(t; \theta)dt \right)$$

and

$$\partial_\theta s(\theta_0, \theta_0) = -\mathbf{E}_{\theta_0} \left( \int_0^1 \frac{\partial_\theta \lambda(t; \theta_0)^{\otimes 2}}{\lambda(t; \theta_0)} dt \right).$$

Then  $-\partial_\theta s(\theta_0, \theta_0) = I(\theta_0)$  and this matrix is semi-definite negative in a neighborhood of  $\theta_0$ , by assumption.

It is sufficient to show that  $\theta_0$  is a separated zero of the limit  $S(\theta)$  of the

normalized score function, in a sense that

$$\sup_{\|\theta - \theta_0\| > \varepsilon} \|S(\theta)\| > \|S(\theta_0)\| = 0 \quad \text{for all } \varepsilon > 0.$$

To prove the uniform convergence

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n} S_n(\theta) - s(\theta, \theta_0) \right\| \xrightarrow{\mathbf{P}} 0$$

it is sufficient to prove the convergence for every  $\theta \in \Theta$  and conclude with the compactness of  $\Theta$ . Then, for every  $\theta$ , by the law of large numbers,  $\frac{1}{n} S_n(\theta) - s(\theta, \theta_0) \rightarrow 0$  in probability. (See Chen & Hall (2016) and Z-estimation theory in Van der Vaart (2000).)

. For the convergence in distribution, we consider a Taylor expansion around  $\theta_0$  :

$$0 = \frac{1}{\sqrt{n}} S_n(\hat{\theta}) = \frac{1}{\sqrt{n}} S_n(\theta_0) + \frac{1}{n} \partial_\theta S_n(\theta^*) \times \sqrt{n}(\hat{\theta} - \theta_0)$$

Assuming that  $\partial_\theta S_n(\theta^*)$  is invertible,

$$\sqrt{n}(\hat{\theta} - \theta_0) = - \left( \frac{1}{n} \partial_\theta S_n(\theta^*) \right)^{-1} \frac{1}{\sqrt{n}} S_n(\theta_0)$$

Hence, by the multivariate CLT, as the processes are i.i.d.,

$$\frac{1}{\sqrt{n}} S_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0))$$

575 And by continuity with respect to  $\theta$ ,  $\frac{1}{n} \partial_\theta S_n(\theta^*) \rightarrow -I(\theta_0)$  in probability. The conclusion comes from Slutsky's lemma.  $\square$

*Proof of Proposition 6.* We follow the idea of Chen & Hall (2016). Let  $\{t_i\}$  the ordered collection of jumps of  $\mathcal{N}$ . Then

$$\hat{h}(t) = \frac{1}{nb_n} \sum_{t_i} K \left( \frac{t - t_i}{b_n} \right) = \frac{1}{nb_n} \int_0^1 K \left( \frac{t - u}{b_n} \right) d\mathcal{N}(u)$$

As the processes  $N_i$  are i.i.d.,

$$\begin{aligned}\mathbf{E}(\hat{h}(t)) &= \frac{1}{n} \sum_{i=1}^n \mathbf{E} \left( \int_0^1 \frac{1}{b_n} K \left( \frac{t-u}{b_n} \right) dN_i(u) \right) \\ &= \int_0^1 \frac{1}{b_n} K \left( \frac{t-u}{b_n} \right) h(u) du\end{aligned}$$

and

$$\begin{aligned}\mathbf{V}(\hat{h}(t)) &= \frac{1}{(nb_n)^2} \mathbf{V} \left( \int_0^1 K \left( \frac{t-u}{b_n} \right) \left( \underbrace{dN(u) - nh(u)du}_{=d\mathcal{M}(u)} \right) \right) \\ &= \frac{1}{(nb_n)^2} \times n \mathbf{V} \left( \int_0^1 K \left( \frac{t-u}{b_n} \right) dM(u) \right) \\ &= \frac{1}{nb_n} \int_0^1 \frac{1}{b_n} K \left( \frac{t-u}{b_n} \right)^2 h(u) du\end{aligned}$$

where the last equality comes from the Itô's isometry. Then,

$$\begin{aligned}\int_0^1 \frac{1}{b_n} K \left( \frac{t-u}{b_n} \right)^2 h(u) du &= \int_{\frac{t-1}{b_n}}^{\frac{t}{b_n}} K(v)^2 h(t - b_n v) dv \\ &\rightarrow h(t) \int_{-\infty}^{+\infty} K(v)^2 dv \quad \text{as } b_n \rightarrow 0\end{aligned}$$

Hence, there exists a positive constant  $C_1$  such that

$$\mathbf{V}(\hat{h}(t)) \leq \frac{C_1}{nb_n}$$

Then, the bias-variance decomposition for the mean-square error (MSE) is :

$$\mathbf{E} \left( \left( \hat{h}(t) - h(t) \right)^2 \right) = \mathbf{V}(\hat{h}(t)) + \underbrace{\left( \mathbf{E}(\hat{h}(t)) - h(t) \right)^2}_{\text{bias}^2}$$

and, for a given  $t \in (0, 1)$ , the bias term is :

$$\mathbf{E}(\hat{h}(t)) - h(t) = \int_{\frac{t-1}{b_n}}^{\frac{t}{b_n}} K(v) h(t - b_n v) dv - h(t)$$

If  $b_n \rightarrow 0$ ,  $\frac{t}{b_n} \rightarrow +\infty$  and  $\frac{t-1}{b_n} \rightarrow -\infty$ . Assuming that the support of  $K$  is  $[-1, 1]$  without loss of generality, for a large enough  $n$ ,

$$\mathbf{E}(\hat{h}(t)) - h(t) = \int K(v) \{h(t - b_n v) - h(t)\} dv$$

Using a Taylor expansion, there exists  $\tau \in (0, 1)$  such that

$$\mathbf{E}(\hat{h}(t)) - h(t) = \int K(v) \left\{ -b_n v h'(t) + b_n^2 \frac{v^2}{2} h''(t - \tau b_n v) \right\} dv$$

As  $K$  is symmetric,  $\int v K(v) dv = 0$  and there exists a positive constant  $C_2$  such that

$$(\mathbf{E}(\hat{h}(t)) - h(t))^2 \leq C_2 b_n^4$$

□

*Proof of Lemma 4..* Under  $\mathcal{H}_0$ , by triangular inequality,

$$(1) = \mathbf{P}(|\hat{h}((j+1)\Delta) - \hat{h}(j\Delta)| \geq \varepsilon) \leq \mathbf{P}((a) + (b) + (c) \geq \varepsilon)$$

with

$$\begin{aligned} (a) &= |\hat{h}((j+1)\Delta) - h((j+1)\Delta)| \\ (b) &= |h((j+1)\Delta) - h(j\Delta)| \\ (c) &= |\hat{h}(j\Delta) - h(j\Delta)| \end{aligned}$$

Then, denote by  $L$  the Lipschitz constant of  $h$  and assume that  $L\Delta < \varepsilon$ . Therefore

$$\begin{aligned} (1) &\leq \mathbf{P}((a) + (c) \geq \varepsilon - L\Delta) \\ &\leq \mathbf{P}((a) \geq \frac{\varepsilon - L\Delta}{2}) + \mathbf{P}((c) \geq \frac{\varepsilon - L\Delta}{2}) \end{aligned}$$

Let  $\tilde{\varepsilon} = \frac{\varepsilon - L\Delta}{2}$ , then by Markov inequality,

$$\mathbf{P}(|\hat{h}(j\Delta) - h(j\Delta)| \geq \tilde{\varepsilon}) \leq \frac{\mathbf{E}((\hat{h}(j\Delta) - h(j\Delta))^2)}{\tilde{\varepsilon}^2}$$

The proof is complete with Proposition 6. □